

## **REMARKS**

Reconsideration of this application, as amended, is respectfully requested.  
Support for the amendments may be found throughout the application as filed.  
No new matter has been added.

### **35 USC 112**

The Examiner has rejected lines 20-21 on page 8 of the disclosure because it contains an embedded hyperlink and/or other form of browser-executable code. The application has been amended to conform with MPEP S. 608.01

### **35 USC 103**

The Examiner has rejected claims 1-7, 14-20 and 27-33 under 35 U.S.C. §103 as being unpatentable over Dean et al., U.S. Patent No. 6,321,220 ("Dean 220' "). The Examiner has rejected also rejected claims 8-12, 21-25 and 24-38 under 35 U.S.C. § 103 as being unpatentable over Dean 220' in view of 'Automatic resource compilation by analyzing hyperlink structure and associated text', Soumen Chakrabarti, IBM, World Wide Web Conference, 1998.

Claims 1, 14 and 27 have been amended to clearly identify the pre-identifying of implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified.

Claims 1-39 are patentable under 35 U.S.C. § 103 in view of the references cited by the Examiner. None of the cited references teach (nor does the Office Action cite any portion which even suggests) the presently claimed feature of expanding each identified core into a full community, the full community being a subset of the pages regarding a particular topic. Moreover, in contrast to applicants, none of the cited references teach (nor does the Office Action cite any portion which even suggests) the pre-identifying of implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified. The pre-identifying of implicitly defined communities is not in response to a user supplied search query. Applicants invention does not utilize a search query. Applicants invention is actually a broad data mining query of the web graph. The broad data mining query identifies implicitly defined communities which were previously undiscovered.

The Examiner concedes at line 11 of page 3 that Dean 220' does not disclose "expanding each identified core into a full community." Dean 220' merely describes a method and apparatus for preventing topic drift in queries in hyperlinked environments. This is in contrast to the pre-identifying of implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified. In the scheme described by Dean 220', for example in flow chart 200 (see Fig. 2 and col. 5-7) a user inputs a query at a search engine. In response to the input query, Dean 220' retrieves matching URLs, prunes those

which are not on topic and ranks the remaining nodes. In short, Dean 220' merely addresses the providing of more relevant search results in response to a search query input by a user at a search engine. Dean 220' addresses the issue of topic drift at ll. 22-45 of col. 3:

When a user wants to find web pages related to a particular topic, the user enters a query representing that topic into a search engine. The search engine finds a result set containing a list of web pages relating that topic. Using an algorithm like Kleinberg's algorithm, this result set is expanded to include other pages that are at a predetermined distance from the pages in the original result set. However, the content of these new pages might not be on the same topic as the original query. If pages that are not on the original query are ranked highly, then this is called "topic drift."

Topic drift may occur when using connectivity information to enlarge the size of an initial result set to include other pages that are reachable within a few links of the initial result set because pages that are one or two links away do not always match the given query. Topic drift also may occur as a result of the existence of many mutually reinforcing pages in the result set, for example if the hub and authority pages point to each other.

Thus, a need exists for a method of preventing topic drift in hyperlinked environments when an initial result set is enlarged to include pages that may better match a given user query.

Thus, Dean 220' does not teach or suggest a scheme for pre-identifying implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified. Moreover, Dean 220' does not teach or suggest expanding each identified core into a full community, the full community being a subset of the pages regarding a particular topic.

Even adding the teachings of Chakrabarti does not render the present invention obvious. Charkrabarti describes a scheme wherein automatic resource

compiler which, given a topic that is broad and well-represented on the web, will seek out and return a list of Web resources that it considers the most authoritative for that topic. Chakrabarti attempts to provide a scheme which addresses lack of automation in the compilation of authoritative resources pertaining to a given topic, where that topic is broad and well-represented on the web. Chakrabarti utilized a combination of text and link analysis for distilling authoritative web resources. However, Chakrabarti does not teach or suggest a scheme for pre-identifying implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified. Moreover, Chakrabarti does not teach or suggest expanding each identified core into a full community, the full community being a subset of the pages regarding a particular topic. Thus, even if the scheme described in Chakrabarti were somehow incorporated into Dean 220', one would still not arrive at the claimed invention. Chakrabarti clearly fails to cure the deficiencies noted with respect to Dean 220', and, therefore, the claims are patentable over the combination of Dean 220' and Chakrabarti.

Baclawski, U.S. Patent No. 6,505,191 ("Baclawski") fails to cure these deficiencies. Baclawski merely describes an indexing and search engine for extraction of information based on the content of information objects in a database as well as links between information objects. Baclawski also supports queries directed at retrieving information with respect to either outgoing or incoming links, or both. For example, Baclawski can be implemented to determine all the pages that refer to one's own home page. However, Baclawski

does not teach or suggest a scheme for pre-identifying implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified. In addition, Baclawski does not teach or suggest expanding each identified core into a full community, the full community being a subset of the pages regarding a particular topic. Baclawski clearly fails to cure the deficiencies noted with respect to Dean 220' and Chakrabarti, and, therefore, the claims are patentable over the combination of Dean 220', Chakrabarti and Baclawski.

Dean, U.S. Patent No. 6,138,113 ("Dean 113' ") fails to cure these deficiencies. Dean 113' merely provides a scheme for identifying near duplicate pages in a hyperlinked database. In Dean 113', a first and second page are selected for a near duplicate determination. For each page, the number of outgoing links is counted. Pages are marked as near duplicates based on the number of common outgoing links between the two pages. Dean 113' is limited to finding near duplicate pages. However, Dean 113' does not teach or suggest a scheme for pre-identifying implicitly defined communities including groups of pages of common interest, from a collection of hyper-linked pages, wherein the communities have not been previously identified. In addition, Dean 113' does not teach or suggest expanding each identified core into a full community, the full community being a subset of the pages regarding a particular topic. Accordingly, the claims are patentable over the combination of Dean 220', Chakrabarti, Baclawski and Dean 113'.

The Examiner's assertion of obviousness is also suspect. The Examiner asserts at ll. 11-15 on page 3 of the Office Action that:

“Dean does not explicitly disclose expanding each identified core into a full community; however, It would have been obvious to one of ordinary skill in the art at the time the invention was made to have modified the hyperlink methods of Dean and utilized the hyperlink pointing methods as disclosed on column 2, lines 50-54, for providing the user an added benefit of efficient common interest group identifications and expanding the result set taught by Dean.”

There is no teaching, suggestion or motivation to implicitly or explicitly to make the modification suggested by the Examiner. Moreover, there is absolutely no explanation as to how such a modification would produce applicant's invention as claimed.

The Examiner asserts Obviousness in light of lines 50-54 of column 2 of Dean 220', the background portion of Dean in combination with “the hyperlink methods of Dean 220', without identifying what and where “the hyperlink methods” are.

Obviousness can only be established by combining or modifying the teachings of the prior art to produce the claimed invention in which there is some teaching, suggestion, or motivation to do so found either in the references themselves or in the knowledge generally available to one of ordinary skill in the art. *In re Fine*, 837 F.2d 1071 (Fed. Cir. 1988). The Office Action indicates that the references cannot be argued individually when cited in combination, but fails to recognize that such combinations are themselves improper when no motivation for the combination is shown. Indeed, rather than show any reasons

for the recited combinations, it appears the teachings of the present application have been used as a blueprint to gather together and assemble various components of the prior art in the manner contemplated by the present applicant. This approach is a classic example of the use of hindsight reconstruction and cannot properly be used as grounds for rejecting the present claims.

The U.S. Court of Appeals for the Federal Circuit has strongly criticized such use of hindsight by specifically indicating that when an obviousness determination is made based upon a combination of references, even a patent examiner "must show reasons that the skilled artisan, confronted with the same problems as the inventor *and with no knowledge of the claimed invention*, would select the elements from the cited prior art references for combination in the manner claimed." *In re Rouffet*, 149 F.3d 1350, 1357 (Fed. Cir. 1998) (Emphasis added). The Examiner merely arguing in his Office Action of July 7, 2003 that the claimed invention would be obvious to one of ordinary skill in the art based on the combination of the references (e.g., Dean 220' background and Dean 220' specification) is utterly inadequate. *Rouffet*, at 1357. Instead, a motivation, either from the references themselves or the knowledge of those of ordinary skill in the art, for the combination being relied upon needs to be shown. *Rouffet*, at 1357.

In the present case, no such motivation has been shown. Instead, the Examiner attempts to deconstruct the subject matter of the claims of the present application into its constituent components. He further states where each such component may be found in one of the cited references and then concludes that

it would have been obvious to combine the references to arrive at the claimed invention. This bare bones analysis is not sufficient to support a determination of obviousness of the present application. The burden is on the Examiner to show *why* one skilled in the art is so motivated as to come up with the combination being relied upon. *Rouffet*, at 1357-1358 (“If such a rote invocation could suffice to supply a motivation to combine, the more sophisticated scientific fields would rarely, if ever, experience a patentable technical advance. Instead, in complex scientific fields [an infringer or the Patent Office] could routinely identify the prior art elements in an application, invoke the lofty level of skill, and rest its case for [obviousness]. To counter this potential weakness in the obviousness construct, the suggestion to combine requirement stands as a critical safeguard against hindsight analysis and rote application of the legal test for obviousness.”).

Moreover, the Examiner is reminded that to establish a *prima facie* case of obviousness, three basic criteria must be met. First, there must be some suggestion or motivation, either in the references themselves or in the knowledge generally available to one of ordinary skill in the art, to modify the reference or to combine reference teachings. Second, there must be a reasonable expectation of success. Finally, the prior art reference (or references when combined) must teach or suggest all the claim limitations. The teachings or suggestion to make the claimed combination and the reasonable expectation of success must both be found in the prior art, and not based on applicant’s disclosure. *In re Vaeck*, 947 F.2d 488, 20 USPQ2d 1438 (Fed. Cir. 1991). See MPEP S. 2143 – S. 2143.03 for decisions pertinent to each of these criteria.

The initial burden is on the examiner to provide some suggestion of the desirability of doing what the inventor has done. "To support the conclusion that the claimed invention is directed to obvious subject matter, either the references must expressly or impliedly suggest the claimed invention or the examiner must present a convincing line of reasoning as to why the artisan would have found the claimed invention to have been obvious in light of the teachings of the references." *Ex parte Clapp*, 227 USPQ 972, 973 (Bd. Pat. App. & Inter. 1985). See MPEP § 2144 - § 2144.09 for examples of reasoning supporting obviousness rejections.

When the motivation to combine the teachings of the references is not immediately apparent, it is the duty of the examiner to explain why the combination of the teachings is proper. *Ex parte Skinner*, 2 USPQ2d 1788 (Bd. Pat. App. & Inter. 1986). A statement of a rejection that includes a large number of rejections must explain with reasonable specificity at least one rejection, otherwise the examiner procedurally fails to establish a *prima facie* case of obviousness. *Ex parte Blanc*, 13 USPQ2d 1383 (Bd. Pat. App. & Inter. 1989) (Rejection based on nine references which included at least 40 prior art rejections without explaining any one rejection with reasonable specificity was reversed as procedurally failing to establish a *prima facie* case of obviousness.).

If the examiner determines there is factual support for rejecting the claimed invention under 35 U.S.C. 103, the examiner must then consider any evidence supporting the patentability of the claimed invention, such as any evidence in the specification or any other evidence submitted by the applicant.

The ultimate determination of patentability is based on the entire record, by a preponderance of evidence, with due consideration to the persuasiveness of any arguments and any secondary evidence. *In re Oetiker*, 977 F.2d 1443, 24 USPQ2d 1443 (Fed. Cir. 1992). The legal standard of "a preponderance of evidence" requires the evidence to be more convincing than the evidence which is offered in opposition to it. With regard to rejections under 35 U.S.C. 103, the examiner must provide evidence which as a whole shows that the legal determination sought to be proved (i.e., the reference teachings establish a *prima facie* case of obviousness) is more probable than not.

When an applicant submits evidence, whether in the specification as originally filed or in reply to a rejection, the examiner must reconsider the patentability of the claimed invention. The decision on patentability must be made based upon consideration of all the evidence, including the evidence submitted by the examiner and the evidence submitted by the applicant. A decision to make or maintain a rejection in the face of all the evidence must show that it was based on the totality of the evidence. Facts established by rebuttal evidence must be evaluated along with the facts on which the conclusion of obviousness was reached, not against the conclusion itself. *In re Eli Lilly & Co.*, 902 F.2d 943, 14 USPQ2d 1741 (Fed. Cir. 1990).

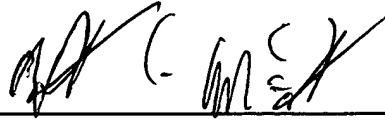
Accordingly, the present rejections under 35 U.S.C. §103(a) should be removed.

If there are any additional charges, please charge Deposit Account No. .

Respectfully submitted,

IBM CORPORATION

Date: 10/7, 2003



---

Mark C. McCabe  
Reg. No. 53,291

IBM Corporation  
Almaden Research Center  
Intellectual Property Law C4TA/J2B  
San Jose, CA 95120-6099  
(408) 927-3380

## **MARKED-UP VERSION OF THE SPECIFICATION**

The following provides a marked-up version of the amendments made to the specification:

**On page 1, replace paragraph one of the “Background of the Invention” section with the following replacement section with the following paragraph:**

The World-wide Web has several thousand well-known, explicitly-defined communities, i.e., groups of individuals who share a common interest, together with the Web pages most popular amongst them. Consider for instance, the community of Web users interested in Porsche Boxster cars. Indeed, there are several explicitly-gathered resource collections, such as those listed under the category of “Recreation: Automotive: Makes and Models: Porsche: Boxster” at the Yahoo Web site (~~www.yahoo.com~~), (yahoo.com) which are devoted to the Boxster. Most of these communities manifest themselves as news groups, Web rings, or as resource collections in directories such as Yahoo! and Infoseek, and as homesteads on Geocities. Other examples include popular topics such as “Major League Baseball,” or the somewhat less visible community of “Prepaid phone card collectors”. The explicit nature of these communities makes them easy to find. It is simply a matter of visiting the appropriate portal or news groups.

**On page 8 please replace the second full paragraph under the Strongly-connected bipartite subgraphs and cores section with the following replacement paragraph:**

Linkage between the related pages can nevertheless be established by a different phenomenon that one can observe on the Web: pages focusing on the same theme frequently contain hyperlinks to the same pages. For instance, as of 12/1/98 the sites [www.swim.org/church.html](http://www.swim.org/church.html), [www.kcm.co.kr/search/church/korea.html](http://www.kcm.co.kr/search/church/korea.html), and [www.cyberkorean.com/church](http://www.cyberkorean.com/church) [swim.org/church.html](http://swim.org/church.html), [kcm.co.kr/search/html](http://kcm.co.kr/search/html), [korea.html](http://korea.html), and [cyberkorean.com/church](http://cyberkorean.com/church) all contain links to numerous Korean churches. This phenomenon is referred to as co-citation, which originated in the bibliometrics literature. See, for instance, Bibliometrics, Annual Review of Information Science and Technology, volume 24, pages 119-186, Elsevier, Amsterdam, 1989. Co-citation suggests that related pages are frequently referenced together. This is even more true in the Web world where linking is not only indicative of good academic discourse, but the essential element that distinguishes the Web as a corpus from other text corpora. For example, the corporate home pages of AT&T and Sprint typically do not reference each other. On the other hand, these pages are very frequently "co-cited". Co-citation is not just a characteristic of well-developed and explicitly-known communities (such as the ones listed above) but an early indicator of newly emerging communities. In other words, the structure of such co-citation in the Web graph can be exploited to extract all communities that have taken shape on the Web, even before the participants have realized that they have formed a community through their co-citation.

**On page 11-12 please replace first paragraph of the "(c) In-degree distribution" section with the following replacement paragraph:**

The first approach to trimming down the resulting data came from an analysis of the in-degrees of Web pages. The distribution of page in-degrees has a remarkably simple rule, as can be seen in the chart of FIG. 4. This chart includes pages that have in-degree at most 410. For any integer  $k$  larger than 410, the chance that a page has in-degree  $k$  is less than 1 in a million. These unusually popular pages (~~e.g., [www.yahoo.com](http://www.yahoo.com)~~) (e.g., [yahoo.com](http://yahoo.com)) with many

potential fans pointing to them have been excluded. The chart suggests a simple relation between in-degree values and their probability densities. Indeed, as can be seen from the remarkably linear log-log plot, the slope of the curve is close to  $\frac{1}{2}$ . This leads to the following empirical fact: the probability that a page has in-degree  $i$  is roughly  $1/i^2$ .

**On page 13 please replace paragraph one of the section entitled “Trawling” with the following replacement paragraph:**

Thus far, several preliminary processing steps on the data have been described, along with some interesting phenomena on degree distributions on the Web graph. The trawling of this "cleaned up" data for communities is now described in detail. The test data still has over 2 million potential fans remaining, with over 60 million links to over 20 million potential centers. Since there are still several million potential fans, it is not practical to enumerate the communities in the form "for all subsets of  $i$  potential fans, and for all subsets of  $j$  potential centers, check if a core is induced". A number of additional pruning steps are therefore necessary to eliminate much of this data, while retaining the property that the eliminated nodes and links cannot be part of any core that is not explicitly identified and output before they are pruned. After the data is reduced by another order of magnitude in this fashion, enumeration of the communities may begin.

**On page 18-19 please replace the second paragraphs of the section entitled “c) Core generation and filtering” with the following:**

Next, nepotistic cores are removed. A nepotistic core is one where some of the fans in the core come from the same Web site. The underlying principle is that if many of the fans in a core come from the same Web site, this may be an artificially established community serving the ends (very likely commercial) of a single entity, rather than a spontaneously-emerging Web community. For this purpose, the following definition of "same Web site" is used. If the site contains

at most three fields, for instance, yahoo.com, or ~~www.ibm.com~~ ibm.com then the site is left as is. If the site has more than three fields, as in ~~www3.yahoo.co.uk~~ yahoo.co.uk, then the first field is dropped. The last column of Table 1 represents the number of non-nepotistic cores. As can be seen, the number of nepotistic cores is significant, but not overwhelming. About half the cores pass the nepotism test.

**On page 21 please replace the section entitled “Communities” with the following replacement paragraph:**

Next, the communities themselves were studied. The following two examples give a sense of the communities that were identified. The first one deals with Japanese pop singer Hekiru Shiina, which has the following fans:

[awa.a-web.co.jp/~buglin/shiina/link.html](http://awa.a-web.co.jp/~buglin/shiina/link.html)

[hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html](http://hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html)

[noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html](http://noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html)

The next example deals with Australian fire brigade services with the following fans:

[maya.eagles.bbs.net.au/~mp/aussie.html](http://maya.eagles.bbs.net.au/~mp/aussie.html)

[homepage.midusa.net/~timcorny/intrnatl.html](http://homepage.midusa.net/~timcorny/intrnatl.html)

[fsinfo.cs.uni-sb.de/~pahu/links\\_australien.html](http://fsinfo.cs.uni-sb.de/~pahu/links_australien.html)

~~http://awa.a-web.co.jp/~buglin/shiina/link.html~~

~~http://hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html~~

~~ml~~

~~http://noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html~~

~~The next example deals with Australian fire brigade services with the following fans:~~

~~http://maya.eagles.bbs.net.au/~mp/aussie.html~~

~~http://homepage.midusa.net/~timcorny/intrnatl.html~~

~~http://fsinfo.cs.uni-sb.de/~pahu/links\_australien.html~~

## **MARKED-UP VERSION OF THE SPECIFICATION**

The following provides a marked-up version of the amendments made to the specification:

**On page 1, replace paragraph one of the “Background of the Invention” section with the following replacement section with the following paragraph:**

The World-wide Web has several thousand well-known, explicitly-defined communities, i.e., groups of individuals who share a common interest, together with the Web pages most popular amongst them. Consider for instance, the community of Web users interested in Porsche Boxster cars. Indeed, there are several explicitly-gathered resource collections, such as those listed under the category of “Recreation: Automotive: Makes and Models: Porsche: Boxster” at the Yahoo Web site (~~www.yahoo.com~~), (yahoo.com) which are devoted to the Boxster. Most of these communities manifest themselves as news groups, Web rings, or as resource collections in directories such as Yahoo! and Infoseek, and as homesteads on Geocities. Other examples include popular topics such as “Major League Baseball,” or the somewhat less visible community of “Prepaid phone card collectors”. The explicit nature of these communities makes them easy to find. It is simply a matter of visiting the appropriate portal or news groups.

**On page 8 please replace the second full paragraph under the Strongly-connected bipartit subgraphs and cores section with the following replacement paragraph:**

Linkage between the related pages can nevertheless be established by a different phenomenon that one can observe on the Web: pages focusing on the same theme frequently contain hyperlinks to the same pages. For instance, as of 12/1/98 the sites [www.swim.org/church.html](http://www.swim.org/church.html), [www.kcm.co.kr/search/church/korea.html](http://www.kcm.co.kr/search/church/korea.html), and [www.cyberkorean.com/church](http://www.cyberkorean.com/church) [swim.org/church.html](http://www.swim.org/church.html), [kcm.co.kr/search/html](http://www.kcm.co.kr/search/html), [korea.html](http://www.kcm.co.kr/search/html), and [cyberkorean.com/church](http://www.cyberkorean.com/church) all contain links to numerous Korean churches. This phenomenon is referred to as co-citation, which originated in the bibliometrics literature. See, for instance, Bibliometrics, Annual Review of Information Science and Technology, volume 24, pages 119-186, Elsevier, Amsterdam, 1989. Co-citation suggests that related pages are frequently referenced together. This is even more true in the Web world where linking is not only indicative of good academic discourse, but the essential element that distinguishes the Web as a corpus from other text corpora. For example, the corporate home pages of AT&T and Sprint typically do not reference each other. On the other hand, these pages are very frequently "co-cited". Co-citation is not just a characteristic of well-developed and explicitly-known communities (such as the ones listed above) but an early indicator of newly emerging communities. In other words, the structure of such co-citation in the Web graph can be exploited to extract all communities that have taken shape on the Web, even before the participants have realized that they have formed a community through their co-citation.

**On page 11-12 please replace first paragraph of the “(c) In-degree distribution” section with the following replacement paragraph:**

The first approach to trimming down the resulting data came from an analysis of the in-degrees of Web pages. The distribution of page in-degrees has a remarkably simple rule, as can be seen in the chart of FIG. 4. This chart includes pages that have in-degree at most 410. For any integer  $k$  larger than 410, the chance that a page has in-degree  $k$  is less than 1 in a million. These unusually popular pages (~~e.g., [www.yahoo.com](http://www.yahoo.com)~~) (e.g., [yahoo.com](http://yahoo.com)) with many

potential fans pointing to them have been excluded. The chart suggests a simple relation between in-degree values and their probability densities. Indeed, as can be seen from the remarkably linear log-log plot, the slope of the curve is close to  $\frac{1}{2}$ . This leads to the following empirical fact: the probability that a page has in-degree  $i$  is roughly  $1/i^2$ .

**On page 13 please replace paragraph one of the section entitled "Trawling" with the following replacement paragraph:**

Thus far, several preliminary processing steps on the data have been described, along with some interesting phenomena on degree distributions on the Web graph. The trawling of this "cleaned up" data for communities is now described in detail. The test data still has over 2 million potential fans remaining, with over 60 million links to over 20 million potential centers. Since there are still several million potential fans, it is not practical to enumerate the communities in the form "for all subsets of  $i$  potential fans, and for all subsets of  $j$  potential centers, check if a core is induced". A number of additional pruning steps are therefore necessary to eliminate much of this data, while retaining the property that the eliminated nodes and links cannot be part of any core that is not explicitly identified and output before they are pruned. After the data is reduced by another order of magnitude in this fashion, enumeration of the communities may begin.

**On page 18-19 please replace the second paragraphs of the section entitled "c) Core generation and filtering" with the following:**

Next, nepotistic cores are removed. A nepotistic core is one where some of the fans in the core come from the same Web site. The underlying principle is that if many of the fans in a core come from the same Web site, this may be an artificially established community serving the ends (very likely commercial) of a single entity, rather than a spontaneously-emerging Web community. For this purpose, the following definition of "same Web site" is used. If the site contains

at most three fields, for instance, yahoo.com, or ~~www.ibm.com~~ ibm.com then the site is left as is. If the site has more than three fields, as in ~~www3.yahoo.co.uk~~ yahoo.co.uk, then the first field is dropped. The last column of Table 1 represents the number of non-nepotistic cores. As can be seen, the number of nepotistic cores is significant, but not overwhelming. About half the cores pass the nepotism test.

**On page 21 please replace the section entitled “Communities” with the following replacement paragraph:**

Next, the communities themselves were studied. The following two examples give a sense of the communities that were identified. The first one deals with Japanese pop singer Hekiru Shiina, which has the following fans:

[awa.a-web.co.jp/~buglin/shiina/link.html](http://awa.a-web.co.jp/~buglin/shiina/link.html)

[hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html](http://hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html)

[noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html](http://noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html)

The next example deals with Australian fire brigade services with the following fans:

[maya.eagles.bbs.net.au/~mp/aussie.html](http://maya.eagles.bbs.net.au/~mp/aussie.html)

[homepage.midusa.net/~timcorny/intrnatl.html](http://homepage.midusa.net/~timcorny/intrnatl.html)

[fsinfo.cs.uni-sb.de/~pahu/links\\_australien.html](http://fsinfo.cs.uni-sb.de/~pahu/links_australien.html)

~~<http://awa.a-web.co.jp/~buglin/shiina/link.html>~~

~~[http://hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.ht](http://hawk.ise.chuo-u.ac.jp/student/person/tshiozak/hobby/heki/hekilink.html)~~

~~ml~~

~~<http://noah.mtl.t.u-tokyo.ac.jp/~msato/hobby/hekiru.html>~~

~~The next example deals with Australian fire brigade services with the following fans:~~

~~<http://maya.eagles.bbs.net.au/~mp/aussie.html>~~

~~<http://homepage.midusa.net/~timcorny/intrnatl.html>~~

~~[http://fsinfo.cs.uni-sb.de/~pahu/links\\_australien.html](http://fsinfo.cs.uni-sb.de/~pahu/links_australien.html)~~